# Computer Vision

James L. Crowley

M2R MoSIG — Fall Semester
4 oct 2018

Lesson 1

## Visual Perception in Man and Machine

**Lesson Outline**:

# 1 The Physics of Light

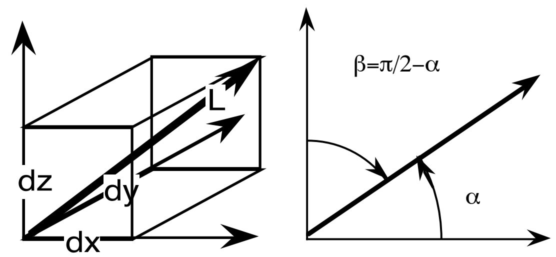
## 1.1 Photons and the Electomagnetic Spectrum

A photon is a resonant electromagnetic oscillation.
The resonance is described by Maxwell's equations.
The magnetic field is strength determined the rate of change of the electric field, and the electric field strength is determined by the rate of change of the magnetic field.

The photon is characterized by
1) a direction of propagation , $\vec{D}$,
2) a direction of oscillation (a polarity), and
3) a wavelength, λ, and its dual a frequency, f : $\lambda = \frac{1}{f}$

The direction of propagation and direction of oscillation can be represented as vectors of cosine angles. For example the direction of propagation can be represented as D.

$$\vec{D} = \begin{pmatrix} \cos(\alpha) \\ \cos(\beta) \\ \cos(\gamma) \end{pmatrix} = \begin{pmatrix} \Delta x / L \\ \Delta y / L \\ \Delta z / L \end{pmatrix}$$




Photon propagation is a probabilistic phenomenon, described by Quantum Chromo-Dynamics. Photons are created and absorbed by abrupt changes in the orbits of electrons. Absorption and creation of photons are probabilistic (non-deterministic) events.

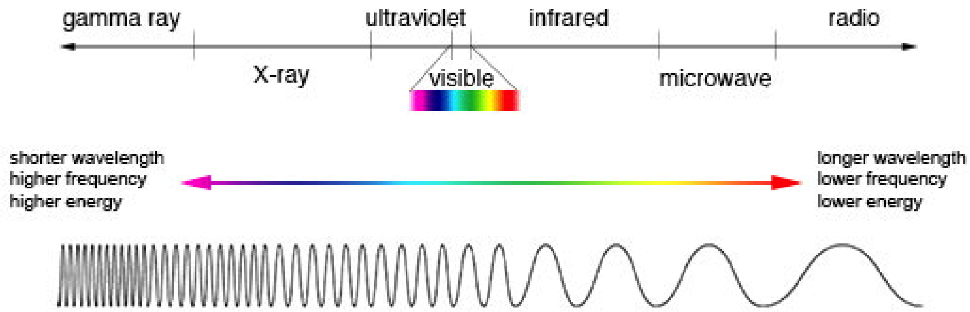
Photons sources generally emit photons over a continuum of directions (a beam) and continuum of wavelengths (a spectrum). The beam intensity is measured in watts, (Joules/sec) and is equivalent to Photons/Meter$^2$.

Radiant flux a measure of the total "amount" of visible light emitted by a source.
The unit of radiant flux can be thought of as a measure of the number of photons of visible light within a beam or angle, or emitted from some source.

Luminous flux (Lumens) is radiant flux adjusted for the wavelengths of light sensed by the retina. 1W = 683 lumens

All electromagnetic energy is propagated as photons. Photons are an oscillation between electric and magnetic potential. The Frequency of the oscillation determines the wavelength of the photon. Energy per unit of spectrum is spectral flux.

The spectrum gives the probability of a photon having a particularly wavelength, S(λ). Visible light is only a very small part of the electromagnetic spectrum.

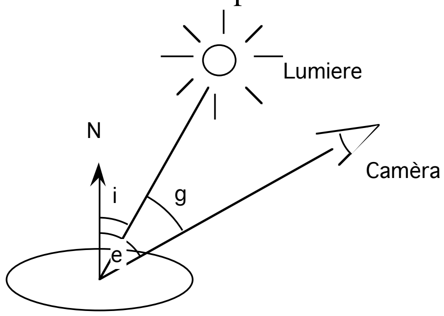



Radio Waves > 1 m (note : 1 Hz = 300,000 km, 100 Mhz = 3 m)
Micro waves: 0.1 cm - 1 m ($10^{-3}$ m to 1 m)
Infrared: 7 x $10^{-7}$ to $10^{-3}$
Visible : 400 nm to 700 nm (4 x $10^{-7}$ m to 7 x $10^{-7}$ m)
Ultraviolet: 10 nm to 400 nm (4x$10^{-7}$ m to $10^{-8}$ m)
X rays: 0.01 nm to 10 nm ($10^{-8}$ m to $10^{-11}$m)
Gamma Rays: ≤ 0.01 nm (≤ $10^{-11}$ m)

## 1.2 Albedo and Reflectance Functions

The albedo of a surface is the ratio of photons emitted over photons received.
Albedo is described by a Reflectance function

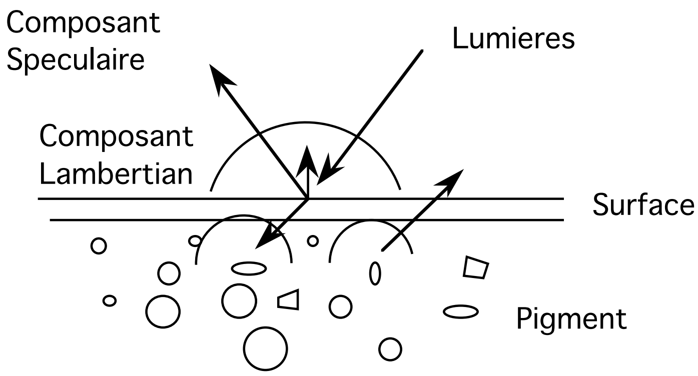
Albedo: $$R(i, e, g, \lambda) = \frac{\text{Number of photons emitted}}{\text{Number of photons received}}$$




The parameters are
i: The incident angle (between the photon source and the normal of the surface).
e: The emittance angle (between the camera and the normal of the surface)
g: The angle between the Camera and the Source.

$\lambda$: The wavelength

For most materials, when photons arrive at a surface, some fraction of the photons are rejected by an interface layer (determined by the wavelength). The remaining photons penetrate the material and are absorbed by molecules near the surface (pigments).




Most reflectance functions can be modeled as a weighted sum of two components: a specular component, $R_S()$ and a Lambertian component, $R_L()$.

$$R(i, e, g, \lambda) = c\, R_S(i, e, g, \lambda) + (1-c)\, R_L(i, \lambda)$$

**Specular Reflection**

The Specular reflection function is given by

$$R_S(i, e, g, \lambda) = \begin{cases} 1 & \text{if } i=e \text{ and } i+e=g \\ 0 & \textit{otherwise} \end{cases}$$

An example of a specular reflector is a mirror.
All (almost all) of the photons are reflected at the interface level with no change in spectrum.

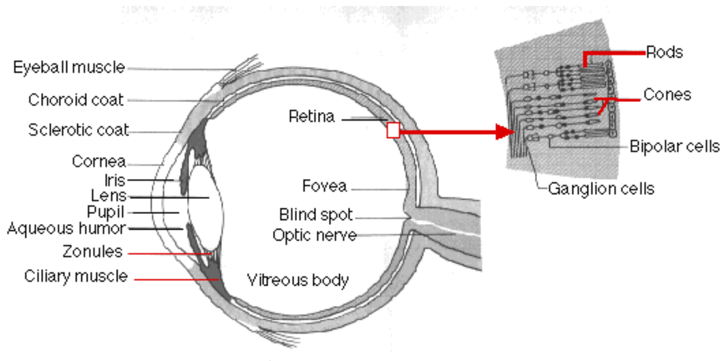
**Lambertian Reflection**

$$R_L(i, \lambda) = P(\lambda)\cos(i)$$

Paper and fresh snow are examples of Lambertian reflectors.

# 2 The Human Visual System

## 2.1 The Human Eye

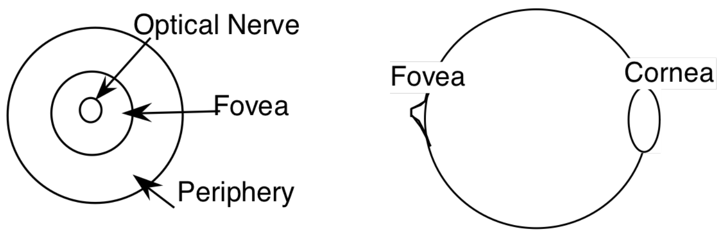



The human eye is a spherical globe filled with transparent liquid.
An opening (iris) allows light to enter and be focused by a lens.
Light arrives at the back of the eye on the Retina.

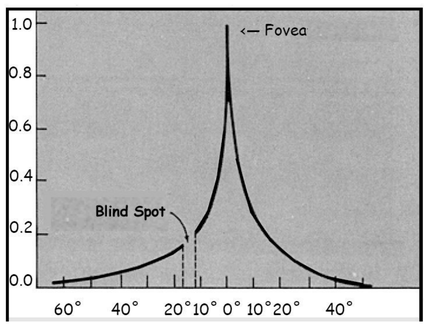
## 2.2 The Retina

The human retina is a tissue composed of a rods, cones and bi-polar cells.
Cones are responsible for daytime vision.
Rods provide night vision.
Bi-polar cells perform initial image processing in the retina.

### Fovea and Peripheral regions




The cones are distributed over a non-uniform region in the back of the eye.
The density of cones decreases exponentially from a central point.
The fovea contains a "hole" where the optic nerve leaves the retina.



The central region of the fovea provides high visual acuity and is used for recognition and depth perception. The peripheral regions have a much lower density of cones, and are used for peripheral vision and to direct eye movements.

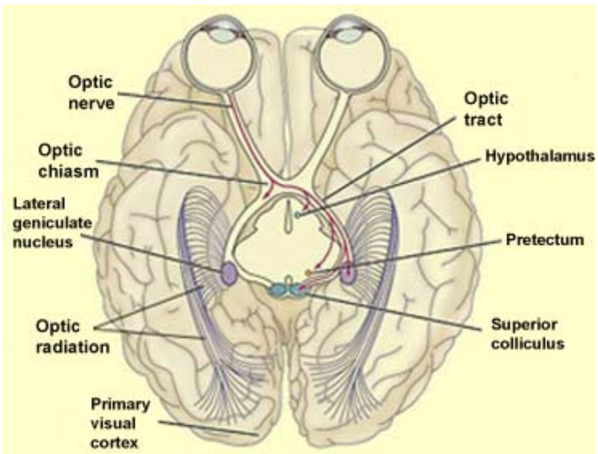
The eye perceives only a small part of the world at any instant.  However, the muscles rotate the eyes at very high rotational accelerations and speeds. The visual system stops processing during such motion, and the observer

The optical nerves leave the retina and are joined at Optic Chiasm.
Nerves then branch off to the Lateral Geniculate Nucleus (LGN) and the Superior Colliculus.

Nerves branch out from the LGN to provide "retinal maps" to the different visual cortexes as well as the Superior Colliculus (SC).

Surprisingly, 80% of the excitation of the LGN comes from the visual cortex!
The LGN acts as a filter for **visual attention**.
In fact, the entire visual system can be seen as succession of filters.

The LGN suppresses neural stimulation that the system is not “attending” to  (looking at).




## 2.3 The Superior Colliculus and the Horopter.

The entire visual system can be seen as succession of filters.
The first visual filter is the iris, that retricts the direction of vision.
This is followed by the rods and cones of the retina, that select photons based on wavelength (or color).
The neurons of the retina are organised as filters that determine an initial point spread function.
The resulting stimulus enter the optical nerves where they are communicated to the LGN and the Superior Colliculus (SC).

The LGN imposes significant top-down fixation, then sends information to the SC.

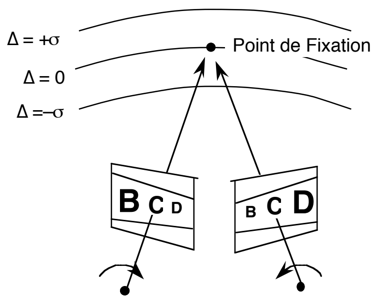
The Superior Colliculus control fixation.

At any instant, the human visual system focuses processing on a small region of 3D space called the Horopter.

The horopter is mathematically defined as the region of space that projects to the same retinal coordinates in both eyes.

**The horopter is the locus of visual fixation in the world**.

The horopter is controlled by the Superior Colliculus, and can move about the scene in incredibly rapid movements (eye scans). Scanning the horopteur allows the cortex to build up a composite model of the external world.




The Superior Colliculus is a Feed-Forward (predictive) control system for binocular fixation. The Superior Colliculus is composed of 7 layers receiving stimulus from the frontal cortex, the lateral and dorsal cortexes, the auditory cortex and the retina.

## 2.4 Vergence and Version

Eye movements can decomposed into "Version" and "Vergence",
Version perceives relative direction in head centered coordinates.
Vergence perceives relative depth.

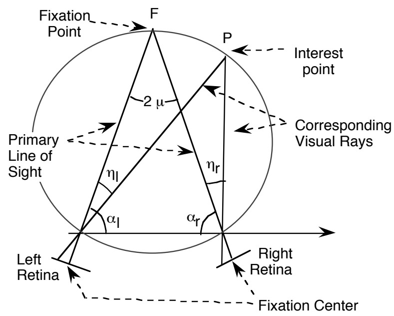
The output of the Superior Colliculus is a neural map that activates Vergence and Version. Depending on the center of neural activation, the eyes are rotated as a sum and a difference.

Version (left-right movement) is the sum of the angles over 2.

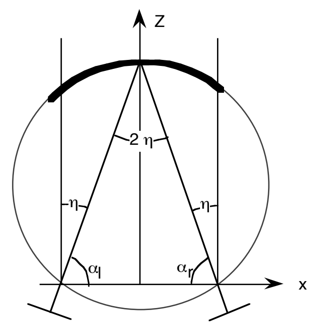
Vergence (in-out movement) is the difference of the angles

Version (angle) is the sum of the eye angles. $\alpha_c = \frac{\alpha_L + \alpha_R}{2}$

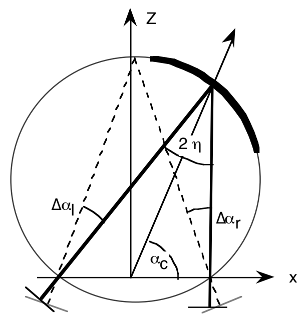
Vergence (depth) is proportional to difference. $\Delta z = \dfrac{\Delta\alpha_L - \Delta\alpha_R}{2}$



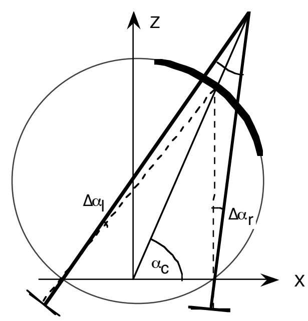

Vergence and version are described by the  Vief-Muller Circle.



Symmetric Vergence

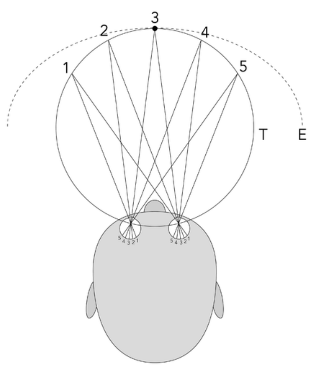


Vergence



Version

Vergence and version are redundantly controlled by retinal matching and by focusing of the lenses in the eyes (accommodation).  Note that human’s lose the ability to accommodate (focus their crystaline lens) around the age of 45 as the lens becomes rigid.

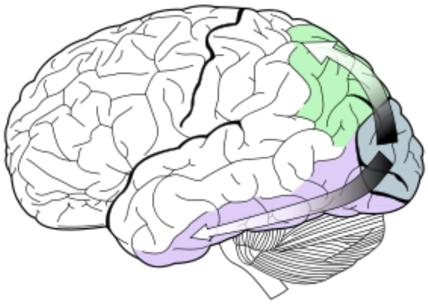



T - Theoretical (mathematical) horopter

E - Empirical (observed) horopter - A twisted cubic surface (larger than T).

## 2.5 The Visual Cortex

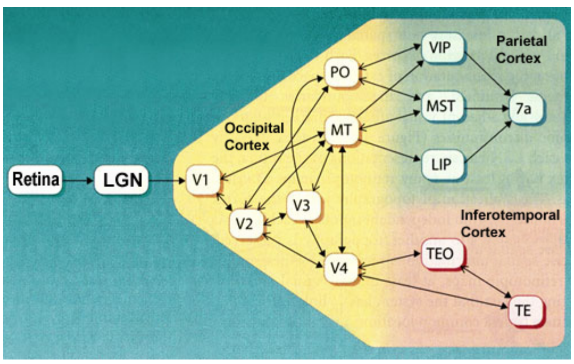
Retinal maps are relayed through the LGN to the primary visual cortex, where they propagate through the Dorsal and Lateral Visual pathways. Only non-suppressed activations are relayed.

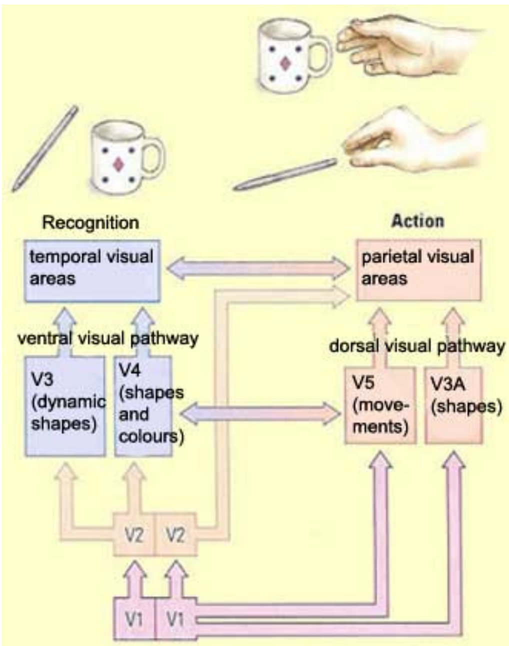


The dorsal visual pathway (green) is the "action pathway".
It controls motor actions. Most of the processing is unconscious.
The dorsal visual pathway makes use of spatial organization (relative 3D position), including depth and direction information from the Superior Colliculus.

The ventral visual pathway (purple) is used in recognizing objects.
It makes use of color and appearance.

These two pathways are divided into a number of interacting subsystems (visual areas).

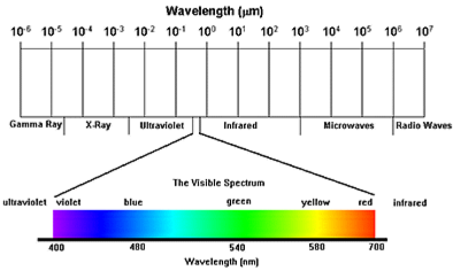


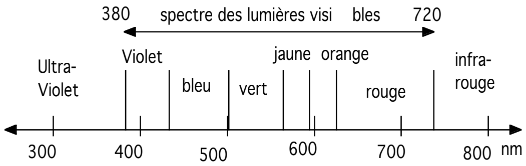




Most human actions require input from both pathways. For example, consider the task of grasping a cup. The brain must recognize and locate the cup, and direct the hand to grasp the cup.

# 3 Color Spaces and Color Models

The human eye is capable of sensing photons with a wavelength between 380 nanometers and 720 nanometers.




Perception of a photon by the retina is also a probabilistic phenomena. (described below).

## 3.1 Color Perception

The human retina is a tissue composed of rods, cones and bi-polar cells.
Cones are responsible for daytime vision.
Bi-polar cells perform initial image processing in the retina.

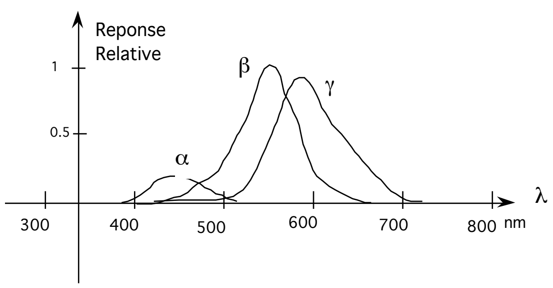
Rods provide night vision. Night vision is achromatique. It does not provide color perception. Night vision is low acuity - Rods are dispersed over the entire retina.

Rods are responsible for perception of very low light levels and provide night vision. Rods employ a very sensitive pigment named "rhodopsin".

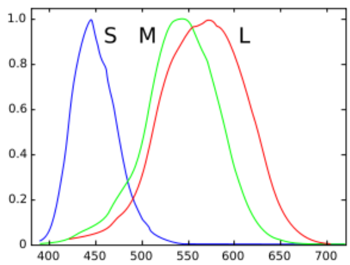
Rodopsin is sensitive to a large part of the visible spectrum of with a maximum sensitivity around 510 nano-meters.

Rhodopsin sensitive to light between 0.1 and 2 lumens, (typical moonlight) but is destroyed by more intense lights.

Rhodopsin can take from 10 to 20 minutes to regenerate.




Relative Sensitivities

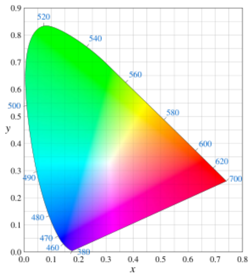



Normalised Sensitivities

Cones provide our chromatique "day vision". Human Cones employ 3 pigments :

cyanolabe α 400–500 nm peak at 420–440 nm
chlorolabe β 450–630 nm peak at 534–545 nm
erythrolabe γ 500–700 nm peak at 564–580 nm

Perception of cyanolabe is low probability, hence poor sensitivity to blue.
Perception of Chlorolabe and erythrolabe are more sensitive.




The three pigments give rise to a color space shown here (CIE model).

Note, these three pigments do NOT map directly to color perception.
Color perception is MUCH more complex, and includes a difficult to model phenomena known as "color constancy". For example, yellow is always yellow, despite changes to the spectrum of an ambiant source

Many color models have been proposed but each has its strengths and weaknesses.

## 3.2 Bayer Matrix Retina

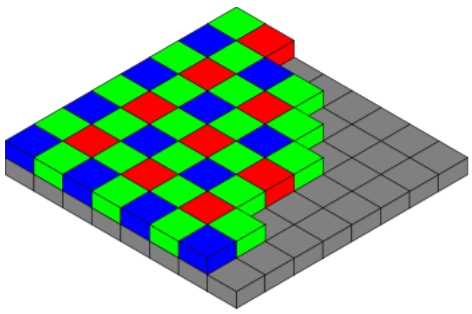
Silicon semiconductors respond to light by emitting photons (Einstein effect), thus generating a charge. A silicon retina is composed of a matrix of individual photocells cells (sensels) that convert photons to positive voltage.

Note that silicon is sensitive to light out into the near infrared ( < 1500 Nm). Color filters are used to limit the spectrum of light reaching each photo-cell.

Most modern digital cameras employ a Bayer Mosaic Retina, named after its inventor, Bryce E. Bayer of Eastman Kodak who patented the design in 1976.

A Bayer filter mosaic is a color filter array (CFA) for arranging RGB color filters on a square grid of photosensors. The filter pattern is 50% green, 25% red and 25% blue, hence is also called RGBG, GRGB, or RGGB.

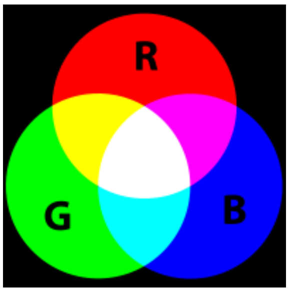
The Bayer mosaic uses twice as many green elements as red or blue to mimic the pigments of the human eye. These elements are referred to as sensor elements, sensels, pixel sensors, or simply pixels;



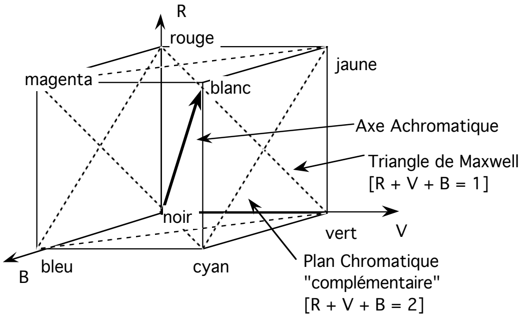
The voltage values on each sensel are converted to numeric values, interpolated and processed to provide image pixels. This step is sometimes called "image reconstruction" in the image processing community, and is generally carried out on the retina. The result is generally an image with colors coded as independent components: RGB.

### 3.3 The RGB Color Model

RGB is one of the oldest color models, originally proposed by Isaac Newton. This is the model used by most color cameras.




The RGB model "pretends" that Red, Green and Blue are orthogonal (independent) axes of a Cartesian space.



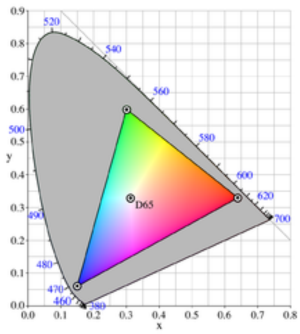

The achromatic axis is R=G=B.
Maxwell's triangle is the surface defined when R+G+B = 1.
A complementary triangle exists when R+G+B = 2.

For printers (subtractive color) this is converted to CMY (Cyan, Magenta, Yellow).

$$\begin{pmatrix} C \\ M \\ Y \end{pmatrix} = \begin{pmatrix} R_{max} \\ G_{max} \\ B_{max} \end{pmatrix} - \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

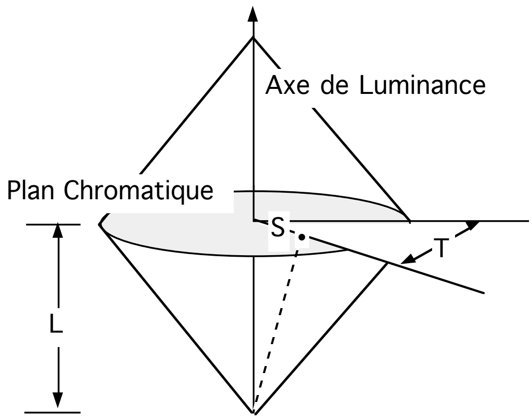
### 3.4 The HLS color model

The RGB model only captures a small part of visible colors:



Painters and artists generally use the HLS: Hue Luminance Saturation model.

HLS is a polar coordinate model for and hue (perceived color) and saturation.
The polar space is placed on a third axis. The size of the disc corresponds to the range of saturation values available.




One (of many possible) mappings from RGB:

Luminance : $L = (R + B + B)$

Saturation : $1 - 3*\min(R, G, B)/L$

Hue : $x = \cos^{-1}\left( \dfrac{\frac{1}{2}(R-G)+(R-B)}{\sqrt{(R-G)^2+(R-B)(G-B)}} \right)$

if B>G then H = x else H = 2π–x.

### 3.5 Color Opponent Model

Color Constancy: The subjective perception of color is independent of the spectrum of the ambient illumination.

Subjective color perception is provided by "Relative" color and not "absolute" number of photons.

This is commonly modeled using a Color Opponent space.

The opponent color theory suggests that there are three opponent channels: red versus green, blue versus yellow, and black versus white (the latter type is achromatic and detects light-dark variation, or luminance).

This can be computed from RGB by the following transformation:

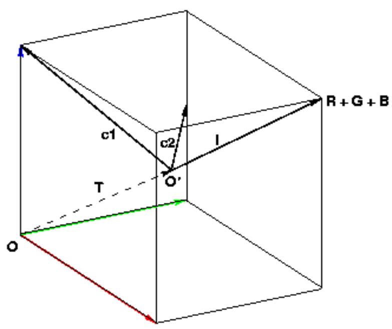
Luminance : $L = R+G+B$

Chrominance: $C1 = (R-G)/2$

$C2 = B - (R+G)/2$

as a matrix :

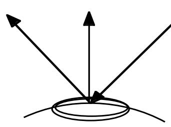
$$\begin{pmatrix} L \\ C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ -0.5 & -0.5 & 1 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$




Such a vector can be "steered" to accommodate changes in ambient illumination.

### 3.6 Separating Specular and Lambertian Reflection.

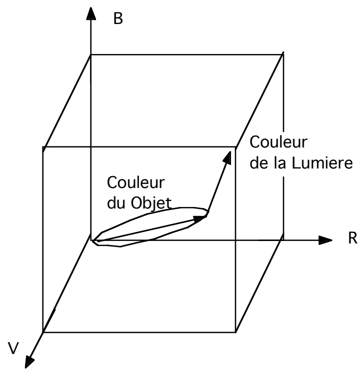
Consider what happens at a specular reflection.



The specularity has the same spectrum as the illumination.
The rest of the object has a spectrum that is the product of illumination and pigments.

This scan be seen in a histogram of color:

$$\forall \vec{C}(i,j) \;:\; H(\vec{C}(i,j)) = H(\vec{C}(i,j)) + 1$$




Two clear axes emerge:
One axis from the origin to the RGB of the product of the illumination and the source.
The other axis towards the RGB representing the illumination.